

A Matrix-Based Method of Moments for Fitting Multivariate Network Meta-Analysis Models with Multiple Outcomes and Random Inconsistency Effects

Dan Jackson^{1,*}, Sylwia Bujkiewicz,² Martin Law¹, Richard D. Riley,³ and Ian R. White^{1,4}

¹MRC Biostatistics Unit, Cambridge, U.K.

²Biostatistics Research Group, Department of Health Sciences, University of Leicester, U.K.

³Centre for Prognosis Research, Research Institute for Primary Care and Health Sciences, University of Keele, U.K.

⁴MRC Clinical Trials Unit at University College London, U.K.

*email: daniel.jackson@mrc-bsu.cam.ac.uk

SUMMARY. Random-effects meta-analyses are very commonly used in medical statistics. Recent methodological developments include multivariate (multiple outcomes) and network (multiple treatments) meta-analysis. Here, we provide a new model and corresponding estimation procedure for multivariate network meta-analysis, so that multiple outcomes and treatments can be included in a single analysis. Our new multivariate model is a direct extension of a univariate model for network meta-analysis that has recently been proposed. We allow two types of unknown variance parameters in our model, which represent between-study heterogeneity and inconsistency. Inconsistency arises when different forms of direct and indirect evidence are not in agreement, even having taken between-study heterogeneity into account. However, the consistency assumption is often assumed in practice and so we also explain how to fit a reduced model which makes this assumption. Our estimation method extends several other commonly used methods for meta-analysis, including the method proposed by DerSimonian and Laird (1986). We investigate the use of our proposed methods in the context of both a simulation study and a real example.

KEY WORDS: Incoherence; Mixed treatment comparisons; Multiple treatments meta-analysis; Random-effects models

1. Introduction

Meta-analysis, the statistical process of pooling the results from separate studies, is commonly used in medical statistics and now requires little introduction. The univariate random-effects model is often used for this purpose. This model has recently been extended to the multivariate (multiple outcomes; Jackson et al., 2011) and network (multiple treatments; Lu and Ades, 2004) meta-analysis settings. In a network meta-analysis, more than two treatments are included in the same analysis. The main advantage of network meta-analysis is that, by using indirect information contained in the network, more precise and coherent inference is possible, especially when direct evidence for particular treatment comparisons is limited. Here, we describe a new model that extends the random-effects modeling framework to the multivariate network meta-analysis setting, so that both multiple outcomes *and* multiple treatments may be included in the same analysis.

Other multivariate extensions of univariate methods for network meta-analysis have previously been proposed. For example, Achana et al. (2014) analyze multiple correlated outcomes in multi-arm studies in public health. Efthimiou et al. (2014) propose a model for the joint modeling of odds ratios on multiple endpoints. Efthimiou et al. (2015) develop another model that is a network extension of an alternative multivariate meta-analytic model that was originally proposed by Riley et al. (2008). A network meta-analysis of multiple

outcomes with individual patient data has also been proposed by Hong et al. (2015) under both contrast-based and arm-based parameterizations, and Hong et al. (2016) develop a Bayesian framework for multivariate network meta-analysis. These multivariate network meta-analysis models are based on the assumption of consistency in the network, extending the approach introduced by Lu and Ades (2004). In contrast to these previously developed methods, the method proposed here relaxes the consistency assumption. This assumption is sometimes found to be false across the entire network (Veroniki et al., 2013). We model the inconsistency using a design-by-treatment interaction, so that different forms of direct and indirect evidence may not agree, even after taking between-study heterogeneity into account. However, we assume that the design-by-interaction terms follow normal distributions, and so conceptualize inconsistency as another source of random variation. This allows us to achieve the dual aim of estimating meaningful treatment effects while also allowing for inconsistency in the network.

Although we allow inconsistency in the network, we propose a relatively simple model. Our preference for a simple model is because the between-study covariance structure is typically hard to identify accurately in multivariate meta-analyses (Jackson et al., 2011) and also because network meta-analysis datasets are usually small (Nikolakopoulou et al., 2014). The new model that we propose for multivariate network meta-

analysis is a direct generalization of the univariate network meta-analysis model proposed by Jackson et al. (2016), which is a particular form of the design-by-treatment interaction model (Higgins et al., 2012). In addition to proposing a new model for multivariate network meta-analysis, we also develop a corresponding new estimation method. This estimation method is based on the method of moments and extends a wide variety of related methods. In particular, we extend the estimation method described by DerSimonian and Laird (1986) by directly extending the matrix based extension of DerSimonian and Laird's estimation method for multivariate meta-analysis (Chen et al., 2012; Jackson et al., 2013). We adopt the usual two-stage approach to meta-analysis, where the estimated study-specific treatment effects (including the within-study covariance matrices) are computed in the first stage. We give some information about how this first stage is performed but our focus is the second stage, where the meta-analysis model is fitted.

The article is set out as follows. In Section 2, we briefly describe the univariate model for network meta-analysis to motivate our new multivariate network meta-analysis model in Section 3. We present our new estimation method in Section 4 and we apply our methods to a real dataset in Section 5. We conclude with a short discussion in Section 6.

2. A Univariate Network Meta-Analysis Model

Here, we describe our univariate modeling framework for network meta-analysis (Jackson et al., 2016; Law et al., 2016). Without loss of generality, we take treatment A as the reference treatment for the network meta-analysis. The other treatments are B, C, *etc.* We take the design d as referring only to the set of treatments compared in a study. For example, if the first design compares treatments A and B only, then $d = 1$ refers to two-arm studies that compare these two treatments. We define t to be the total number of treatments included in the network, and t_d to be the number of treatments included in design d . We define D to be the number of different designs, N_d to be the number of studies of design d , and $N = \sum_{d=1}^D N_d$ to be the total number of studies. We will use the word “contrast” to refer to a particular treatment comparison or effect in a particular study, for example, the “AB contrast” in the first study.

We model the estimated relative treatment effects, rather than the average outcomes in each arm, and so perform contrast based analyses. We define \mathbf{Y}_{di} to be the $c_d \times 1$ column vector of estimated relative treatment effects from the i th study of design d , where $c_d = t_d - 1$. We define $n_d = N_d c_d$ to be the total number of estimated treatment effects that design d contributes to the analysis, and $n = \sum_{d=1}^D n_d$ to be the total number of estimated treatment effects that contribute to the analysis. To specify the outcome data \mathbf{Y}_{di} , we choose a baseline treatment in each design d . The entries of \mathbf{Y}_{di} are then the estimated effects of the other c_d treatments included in design d relative to this baseline treatment. For example, if we take $d = 2$ to indicate the “CDE design” then $c_2 = 2$. Taking C as the baseline treatment for this design, the two entries of the \mathbf{Y}_{2i} vectors are the estimated relative effects of treatment D

compared to C and of treatment E compared to C. For example, the entries of the \mathbf{Y}_{di} could be estimated log-odds ratios or mean differences.

We use normal approximations for the within-study distributions. We define \mathbf{S}_{di} to be the $c_d \times c_d$ within-study covariance matrix corresponding to \mathbf{Y}_{di} . We treat all \mathbf{S}_{di} as fixed and known in analysis. Ignoring the uncertainty in the \mathbf{S}_{di} is acceptable provided that the studies are reasonably large and is conventional in meta-analysis, but this approximation is motivated by pragmatic considerations because this greatly simplifies the modeling. We do not impose any constraints on the form of \mathbf{S}_{di} other than they must be valid covariance matrices. The lead diagonal entries of the \mathbf{S}_{di} are within-study variances that can be calculated using standard methods. Assuming that the studies are composed of independent samples for each treatment, the other entries of the \mathbf{S}_{di} are calculated as the variance of the average outcome (for example the log odds or the sample mean) of the baseline treatment.

We define $\delta_1^{AB}, \delta_1^{AC}, \dots, \delta_1^{AZ}$, where Z is the final treatment in the network, to be treatment effects relative to the reference treatment A, and call them basic parameters (Lu and Ades, 2006). We use the subscript 1 when defining the basic parameters to emphasize that they are treatment effects for the first (and in this section, only) outcome. We define $c = t - 1$ to be the number of basic parameters in the univariate setting. Treatment effects not involving A can be obtained as linear combinations of the basic parameters and are referred to as functional parameters (Lu and Ades, 2006). For example, the average treatment effect of treatment E to treatment C, $\delta_1^{CE} = \delta_1^{AE} - \delta_1^{AC}$, is a functional parameter. We define the $c \times 1$ column vector $\boldsymbol{\delta} = (\delta_1^{AB}, \delta_1^{AC}, \dots, \delta_1^{AZ})^T$ and design specific $c_d \times c$ design matrices $\mathbf{Z}_{(d)}$. We use the subscript (d) in these design matrices to emphasize that they apply to each individual study of design d ; we reserve the subscript d for design matrices that describe regression models for all outcome data from this design. If the i th entry of the \mathbf{Y}_{di} are estimated treatment effects of treatment J relative to the reference treatment A then the i th row of $\mathbf{Z}_{(d)}$ contains a single nonzero entry: 1 in the $(j - 1)$ th column, where j is the position of J in the alphabet. If instead the i th entry of the \mathbf{Y}_{di} are estimated treatment effects of treatment J relative to treatment K, $K \neq A$, then the i th row of $\mathbf{Z}_{(d)}$ contains two nonzero entries: 1 in the $(j - 1)$ th column and -1 in the $(k - 1)$ th column.

Our univariate model for network meta-analysis is

$$\mathbf{Y}_{di} = \mathbf{Z}_{(d)}\boldsymbol{\delta} + \boldsymbol{\Theta}_{di} + \boldsymbol{\Omega}_d + \boldsymbol{\epsilon}_{di} \quad (1)$$

where $\boldsymbol{\Theta}_{di} \sim N(\mathbf{0}, \tau_\beta^2 \mathbf{P}_{c_d})$, $\boldsymbol{\Omega}_d \sim N(\mathbf{0}, \tau_\omega^2 \mathbf{P}_{c_d})$, $\boldsymbol{\epsilon}_{di} \sim N(\mathbf{0}, \mathbf{S}_{di})$, all $\boldsymbol{\Theta}_{di}$, $\boldsymbol{\Omega}_d$, and $\boldsymbol{\epsilon}_{di}$ are independent, and \mathbf{P}_{c_d} is the $c_d \times c_d$ matrix with ones on the leading diagonal and halves elsewhere. We refer to τ_β^2 and τ_ω^2 as the between-study variance, and the inconsistency variance, respectively. The term $\boldsymbol{\Theta}_{di}$ is a study-by-treatment interaction term that models between-study heterogeneity. The model $\boldsymbol{\Theta}_{di} \sim N(\mathbf{0}, \tau_\beta^2 \mathbf{P}_{c_d})$ implies that the heterogeneity variance is the same for all contrasts for every study regardless of whether or not the comparison is relative to the baseline treatment (Lu and Ades, 2004). Other

simple choices of \mathbf{P}_{cd} , such as allowing the off-diagonal entries to differ from 0.5, violate this symmetry between treatments. For example, in the case $d = 2$ indicating the CDE design, the between-study variances for the CD and CE effects in this study are given by the two diagonal entries of $\tau_\beta^2 \mathbf{P}_{cd}$, which are both τ_β^2 . The between-study variance for the effect of E relative to D is $(-1, 1)\tau_\beta^2 \mathbf{P}_{cd}(-1, 1)^T$, which is also τ_β^2 . The $\mathbf{\Omega}_d$ are design-by-treatment interaction terms that model inconsistency in the network. The model $\mathbf{\Omega}_d \sim N(\mathbf{0}, \tau_\omega^2 \mathbf{P}_{cd})$ implies that the inconsistency variance is the same for all contrasts for every design; other simple choices of \mathbf{P}_{cd} violate this symmetry.

To describe all estimates from all studies, we stack the \mathbf{Y}_{di} from the same design to form the $n_d \times 1$ column vector $\mathbf{Y}_d = (\mathbf{Y}_{d1}^T, \dots, \mathbf{Y}_{dN_d}^T)^T$, and we then stack these \mathbf{Y}_d to form the $n \times 1$ column vector $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_D^T)^T$. Jackson et al. (2016) then use three further matrices that we also define here because they will be required to describe the estimation procedure that follows. The matrix \mathbf{M}_1 is defined as a $n \times n$ square matrix where $m_{1ij} = 0$ if the i th and j th entries of \mathbf{Y} , $i, j = 1, \dots, n$, are estimates from different studies; otherwise $m_{1ii} = 1$, and $m_{1ij} = 1/2$ for $i \neq j$. The matrix \mathbf{M}_2 is defined as a $n \times n$ square matrix where $m_{2ij} = 0$ if the i th and j th entries of \mathbf{Y} , $i, j = 1, \dots, n$, are estimates from different designs; otherwise $m_{2ij} = 1$ if the i th and j th entries of \mathbf{Y} are estimates of the same treatment comparison (for example, treatment A compared to treatment B) and $m_{2ij} = 1/2$ if these entries are estimates of different treatment comparisons. The supplementary materials show a concrete example showing how these two matrices are formed. Jackson et al. (2016) also define a $n \times c$ univariate design matrix \mathbf{Z} , where if the i th entry of \mathbf{Y} is an estimated treatment effect of treatment J relative to the reference treatment A then the i th row of \mathbf{Z} contains a single nonzero entry: 1 in the $(j-1)$ th column, where j is the position of J in the alphabet. If instead the i th entry of \mathbf{Y} is an estimated treatment effect of treatment J relative to treatment K, $K \neq A$, then the i th row of \mathbf{Z} contains two nonzero entries: 1 in the $(j-1)$ th column and -1 in the $(k-1)$ th column. Defining $\mathbf{S}_d = \text{diag}(\mathbf{S}_{d1}, \dots, \mathbf{S}_{dN_d})$, and then $\mathbf{S} = \text{diag}(\mathbf{S}_1, \dots, \mathbf{S}_D)$, model (1) can be presented for the entire dataset as

$$\mathbf{Y} \sim N(\mathbf{Z}\boldsymbol{\delta}, \tau_\beta^2 \mathbf{M}_1 + \tau_\omega^2 \mathbf{M}_2 + \mathbf{S}).$$

3. A Multivariate Network Meta-Analysis Model

We now explain how to extend the univariate model in Section 2 to the multivariate setting to handle multiple outcomes. We define p to be the number of outcomes, and so the dimension of the network meta-analysis, so that we now consider the case where $p > 1$. The \mathbf{Y}_{di} are now $pc_d \times 1$ column vectors, where the \mathbf{Y}_{di} contain c_d column vectors of length p . For example, in a $p = 5$ dimensional meta-analysis and continuing with the example where $d = 2$ indicates the CDE design, we have $c_2 = 2$. The \mathbf{Y}_{2i} are then 10×1 column vectors where, taking C as the baseline treatment for this design, the first five entries of the \mathbf{Y}_{2i} are estimated relative treatment effect of D compared to C and the second five entries are the same estimate of E compared to C. We define the $pc \times 1$ column vector $\boldsymbol{\delta} =$

$(\delta_1^{AB}, \delta_1^{AC}, \dots, \delta_1^{AZ}, \delta_2^{AB}, \delta_2^{AC}, \dots, \delta_2^{AZ}, \dots, \delta_p^{AB}, \delta_p^{AC}, \dots, \delta_p^{AZ})^T$, so that this vector contains the basic parameters for each outcome in turn. When $p = 1$ the vector $\boldsymbol{\delta}$ reduces to its definition in the univariate setting, as given in Section 2.

We define $\boldsymbol{\Sigma}_\beta$ and $\boldsymbol{\Sigma}_\omega$ to be $p \times p$ unstructured covariance matrices that are multivariate generalizations of τ_β^2 and τ_ω^2 . These two matrices contain the between-study variances and covariances, and the inconsistency variances and covariances, respectively, for all p outcomes. We refer to $\boldsymbol{\Sigma}_\beta$ and $\boldsymbol{\Sigma}_\omega$ as the between-study covariance matrix, and the inconsistency covariance matrix, respectively. We continue to treat the within-study covariance matrices \mathbf{S}_{di} as if fixed and known in analysis but these are now $pc_d \times pc_d$ matrices. The entries of the \mathbf{S}_{di} matrices that describe the covariance of estimated treatment effects for the same outcome can be obtained as in the univariate setting. However, the other entries of \mathbf{S}_{di} , that describe the covariance between treatment effects for different outcomes, are harder to obtain in practice. A variety of strategies for dealing with this difficulty have been proposed (Jackson et al., 2011; Wei and Higgins, 2013).

3.1. The Proposed Multivariate Model for Network Meta-Analysis

In the multivariate setting, to allow correlations between estimated treatment effects for different outcomes, both within studies and designs, we propose that model (1) is generalized to

$$\mathbf{Y}_{di} = \mathbf{X}_{(d)}\boldsymbol{\delta} + \boldsymbol{\Theta}_{di} + \boldsymbol{\Omega}_d + \boldsymbol{\epsilon}_{di} \quad (2)$$

where $\mathbf{X}_{(d)} = ((\mathbf{I}_p \otimes \mathbf{Z}_{(d)1})^T, \dots, (\mathbf{I}_p \otimes \mathbf{Z}_{(d)c_d})^T)^T$, $\mathbf{Z}_{(d)i}$ is the i th row of $\mathbf{Z}_{(d)}$, $\boldsymbol{\Theta}_{di} \sim N(\mathbf{0}, \mathbf{P}_{cd} \otimes \boldsymbol{\Sigma}_\beta)$, $\boldsymbol{\Omega}_d \sim N(\mathbf{0}, \mathbf{P}_{cd} \otimes \boldsymbol{\Sigma}_\omega)$ and $\boldsymbol{\epsilon}_{di} \sim N(\mathbf{0}, \mathbf{S}_{di})$, where all $\boldsymbol{\Theta}_{di}$, $\boldsymbol{\Omega}_d$, and $\boldsymbol{\epsilon}_{di}$ are independent, and \otimes is the Kronecker product. The random $\boldsymbol{\Theta}_{di}$ and $\boldsymbol{\Omega}_d$ continue to model between-study heterogeneity, and inconsistency, respectively. Recalling that $\boldsymbol{\delta}$ contains the basic parameters for each outcome in turn, the design matrices $\mathbf{X}_{(d)}$ provide the correct linear combinations of basic parameters to describe the mean of all estimated treatment effects in \mathbf{Y}_{di} . Model (2) reduces to model (1) in one-dimension. The definition of \mathbf{P}_{cd} means that $\boldsymbol{\Sigma}_\beta$ and $\boldsymbol{\Sigma}_\omega$ are the between-study covariance matrix, and inconsistency covariance matrix, for all contrasts. We continue define \mathbf{Y} as in the univariate setting, where \mathbf{Y} contains n column vectors of estimated treatment effects that are of length p , so that \mathbf{Y} is a $np \times 1$ column vector in the multivariate setting. We define the multivariate $np \times pc$ design matrix $\mathbf{X} = ((\mathbf{I}_p \otimes \mathbf{Z}_1)^T, \dots, (\mathbf{I}_p \otimes \mathbf{Z}_n)^T)^T$, where \mathbf{Z}_i is the i th row of \mathbf{Z} . Model (2) can be presented for the entire dataset as

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\delta}, \mathbf{M}_1 \otimes \boldsymbol{\Sigma}_\beta + \mathbf{M}_2 \otimes \boldsymbol{\Sigma}_\omega + \mathbf{S}) \quad (3)$$

where we continue to define \mathbf{S} as in the univariate case. Matrices \mathbf{M}_1 and \mathbf{M}_2 are the same as in the univariate setting, and so continue to be $n \times n$ matrices. Model (3) is a linear mixed model for network meta-analysis and is conceptually similar to other models of this type (Piepho et al., 2012). If $\boldsymbol{\Sigma}_\omega = \mathbf{0}$ then all $\boldsymbol{\Omega}_d = \mathbf{0}$ and there is no inconsistency; we refer to this reduced model as the “consistent model.” If both $\boldsymbol{\Sigma}_\beta = \mathbf{0}$ and $\boldsymbol{\Sigma}_\omega = \mathbf{0}$ then all studies estimate the same effects

to within-study sampling error and we refer to this model as the “common-effect and consistent model.”

Missing data (unobserved entries of \mathbf{Y}) are common in applications as not all studies may provide data for all outcomes and contrasts. When there are missing outcome data, the model for the observed data is the marginal model for the observed data implied by (3), where any rows of \mathbf{Y} that contain missing values are discarded. We will use a non-likelihood based approach for making inferences and so assume any data are missing completely at random (Seaman et al., 2013). We define the diagonal $np \times np$ missing indicator matrix \mathbf{R} , where $\mathbf{R}_{ii} = 1$ if \mathbf{Y}_i is observed, $\mathbf{R}_{ii} = 0$ if \mathbf{Y}_i is missing, and $\mathbf{R}_{ij} = 0$ if $i \neq j$.

4. Multivariate Estimation: A New Method of Moments

Our estimation procedure is motivated by the univariate method proposed by DerSimonian and Laird (1986). This was developed in the much simpler setting where each study provides a single estimate Y_i , and where the random-effects model $Y_i \sim N(\delta, \tau^2 + S_i)$ is assumed. This estimation method for τ^2 uses the Q statistic, where $Q = \sum S_i^{-1}(Y_i - \hat{\delta})^2$ and $\hat{\delta} = \sum S_i^{-1}Y_i / \sum S_i^{-1}$ is the pooled estimate under the common-effect model ($\tau^2 = 0$).

Now consider an alternative representation of this Q statistic. Taking $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\mathbf{S} = \text{diag}(S_1, \dots, S_n)$, and $\mathbf{W} = \mathbf{S}^{-1}$ means that $Q = \text{tr}(\mathbf{W}(\mathbf{Y} - \hat{\mathbf{Y}})(\mathbf{Y} - \hat{\mathbf{Y}})^T)$, where $\hat{\mathbf{Y}}$ is obtained under the common-effect model. To obtain a $p \times p$ matrix generalization of Q for multivariate analyses, we replace the trace operator with the block trace operator in this expression (Jackson et al., 2013). The block trace operator is a generalization of the trace that sums over all n of the $p \times p$ matrices along the main block diagonal of an $np \times np$ matrix. This produces a $p \times p$ matrix. In the absence of missing data we can write our multivariate generalization of the Q statistic, $\text{btr}(\mathbf{W}(\mathbf{Y} - \hat{\mathbf{Y}})(\mathbf{Y} - \hat{\mathbf{Y}})^T)$, as a weighted sum of outer products of $p \times 1$ vectors of residuals under the common-effect and consistent model. Hence, the distribution of $\text{btr}(\mathbf{W}(\mathbf{Y} - \hat{\mathbf{Y}})(\mathbf{Y} - \hat{\mathbf{Y}})^T)$ depends directly on the magnitudes of unknown variance components.

4.1. A Q Matrix for Multivariate Network Meta-Analysis

We define a within-study precision matrix \mathbf{W} corresponding to \mathbf{S} . If there are no missing outcome data in \mathbf{Y} then we define $\mathbf{W} = \mathbf{S}^{-1}$, where \mathbf{S} is taken from model (3). If there are missing data in \mathbf{Y} then the entries of \mathbf{W} that correspond to observed data are obtained as the inverse of the corresponding entries of the within-study covariance matrix of reduced dimension (equal to that of the observed data) and the other entries of \mathbf{W} are set to zero. For example, consider the case where \mathbf{Y} is a 6×1 vector but only the second and fifth entries are observed; this corresponds to much less outcome data than would be used in practice but provides an especially simple example. Then we define \mathbf{S}_r , where the subscript r indicates a dimension reduction, as a 2×2 matrix whose entries are the within-study variances and covariances of the two observed entries of \mathbf{Y} . The 6×6 precision matrix \mathbf{W} then has all zero entries in the first, third, fourth, and sixth rows and columns. However, the remaining entries of \mathbf{W} are the entries of the 2×2 matrix \mathbf{S}_r^{-1} ,

so that $\mathbf{W}_{22} = (\mathbf{S}_r^{-1})_{11}$, $\mathbf{W}_{25} = (\mathbf{S}_r^{-1})_{12}$, $\mathbf{W}_{52} = (\mathbf{S}_r^{-1})_{21}$, and $\mathbf{W}_{55} = (\mathbf{S}_r^{-1})_{22}$. We define $\hat{\mathbf{Y}}$ to be the fitted value of \mathbf{Y} under the common-effect and consistent model ($\boldsymbol{\Sigma}_\beta = \boldsymbol{\Sigma}_\omega = \mathbf{0}$), so that $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$ where $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}$. We also define an asymmetric $np \times np$ matrix (Jackson et al., 2013)

$$\mathbf{Q} = \mathbf{W}\{\mathbf{R}(\mathbf{Y} - \hat{\mathbf{Y}})\}\{\mathbf{R}(\mathbf{Y} - \hat{\mathbf{Y}})\}^T = \mathbf{W}(\mathbf{Y} - \hat{\mathbf{Y}})(\mathbf{Y} - \hat{\mathbf{Y}})^T\mathbf{R} \quad (4)$$

Our definitions of \mathbf{W} and \mathbf{R} mean that $\mathbf{WR} = \mathbf{W}$, which results in the simplified version of \mathbf{Q} in (4). From the first form given in (4), we have that the residuals $\mathbf{Y} - \hat{\mathbf{Y}}$ are pre-multiplied by \mathbf{R} , so that any residuals that correspond to missing outcome data do not contribute to \mathbf{Q} . Furthermore, missing outcome data do not contribute to $\hat{\mathbf{Y}}$ because they have no weight under the common-effect and consistent model. Hence, we can impute missing outcome data with any finite value without changing the value of \mathbf{Q} . This is merely a convenient way to handle missing data numerically and has no implications for the statistical modeling.

4.2. Design Specific Q Matrices for Multivariate Network Meta-Analysis

In order to identify the full model, we will require design-specific versions of \mathbf{Q} that only use data from a particular design. As in the univariate setting, we stack the outcome data from design d to form the vector $\mathbf{Y}_d = (\mathbf{Y}_{d1}^T, \dots, \mathbf{Y}_{dN_d}^T)^T$. In the multivariate setting, the vector \mathbf{Y}_d contains n_d estimated effects each of length p , so that \mathbf{Y}_d is now a $pn_d \times 1$ column vector. We define the design specific $n_d \times n_d$ matrix \mathbf{M}_1^d , where $m_{1ij}^d = 0$ if the i th and j th estimated effect (of length p) in \mathbf{Y}_d , $i, j = 1, \dots, n_d$, are from separate studies; otherwise $m_{1ii}^d = 1$ and $m_{1ij}^d = 1/2$ for $i \neq j$. We define the $pn_d \times pc_d$ design matrix \mathbf{X}_d which is obtained by stacking identity matrices of dimension pc_d , where we include one such identity matrix for each study of design d . Hence, $\mathbf{X}_d = \mathbf{1}_{N_d} \otimes \mathbf{I}_{pc_d}$, where $\mathbf{1}_{N_d}$ is the $N_d \times 1$ column vector where every entry is one. We also define the $pc_d \times 1$ column vector $\boldsymbol{\beta}_d = \mathbf{X}_{(d)}\boldsymbol{\delta} + \boldsymbol{\Omega}_d$.

An identifiable design-specific marginal model for outcome data from design d only, that is implied by model (2), is

$$\mathbf{Y}_d \sim N(\mathbf{X}_d\boldsymbol{\beta}_d, \mathbf{M}_1^d \otimes \boldsymbol{\Sigma}_\beta + \mathbf{S}_d) \quad (5)$$

where $\mathbf{S}_d = \text{diag}(\mathbf{S}_{d1}, \dots, \mathbf{S}_{dN_d})$. We can also calculate design specific versions of (4) where we calculate all quantities, including the fitted values, using just the data from studies of design d . We define these $pn_d \times pn_d$ design specific matrices as

$$\mathbf{Q}_d = \mathbf{W}_d(\mathbf{Y}_d - \hat{\mathbf{Y}}_d)(\mathbf{Y}_d - \hat{\mathbf{Y}}_d)^T\mathbf{R}_d \quad (6)$$

where \mathbf{W}_d , \mathbf{R}_d , and $\hat{\mathbf{Y}}_d$ in (6) are defined in the same way as \mathbf{W} , \mathbf{R} and $\hat{\mathbf{Y}}$ in (4) but where only data from design d are used. Hence, \mathbf{R}_d and \mathbf{W}_d are the missing indicator matrix, and the within-study precision matrix, of \mathbf{Y}_d , respectively. We compute $\hat{\mathbf{Y}}_d = \mathbf{H}_d\mathbf{Y}_d$ where $\mathbf{H}_d = \mathbf{X}_d(\mathbf{X}_d^T\mathbf{W}_d\mathbf{X}_d)^{-1}\mathbf{X}_d^T\mathbf{W}_d$. When computing \mathbf{H}_d we take the matrix inverse to be the Moore–Penrose pseudoinverse (Searle, 1971). This is so that

any design-specific regression corresponding to this hat matrix that is not fully identifiable (due to missing outcome data) can still contribute to the estimation. We use model (5) to derive the properties of \mathbf{Q}_d in equation (6).

4.3. The Estimating Equations

We base our estimation on the two $p \times p$ matrices $\text{btr}(\mathbf{Q})$ and $\sum_{d=1}^D \text{btr}(\mathbf{Q}_d)$, where \mathbf{Q} and \mathbf{Q}_d are given in (4) and (6), respectively. Specifically, we match these quantities to their expectations to estimate the unknown variance parameters using the method of moments.

4.3.1. Evaluating $E[\text{btr}(\mathbf{Q})]$ and deriving the first estimating equation. We define $\mathbf{A} = (\mathbf{I}_{np} - \mathbf{H})^T \mathbf{W}$ and $\mathbf{B} = (\mathbf{I}_{np} - \mathbf{H})^T \mathbf{R}$, which are known $np \times np$ matrices. We also divide the matrices \mathbf{A} and \mathbf{B} into n^2 blocks of $p \times p$ matrices, and write $\mathbf{A}_{i,j}$ and $\mathbf{B}_{i,j}$, $i, j = 1, \dots, n$, to mean the i th by j th blocks of \mathbf{A} and \mathbf{B} respectively. Hence, $\mathbf{A}_{i,j}$ and $\mathbf{B}_{i,j}$ are both $p \times p$ matrices. In the supplementary materials, we show that

$$\begin{aligned} E[\text{btr}(\mathbf{Q})] &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n m_{1ij} \mathbf{A}_{k,i} \boldsymbol{\Sigma}_\beta \mathbf{B}_{j,k} \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n m_{2ij} \mathbf{A}_{k,i} \boldsymbol{\Sigma}_\omega \mathbf{B}_{j,k} + \text{btr}(\mathbf{B}). \end{aligned}$$

We apply the $\text{vec}(\cdot)$ operator to both sides of the previous equation and use the identity $\text{vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{X})$ (see Henderson and Searle, 1981), to obtain

$$\text{vec}(E[\text{btr}(\mathbf{Q})]) = \mathbf{C}\text{vec}(\boldsymbol{\Sigma}_\beta) + \mathbf{D}\text{vec}(\boldsymbol{\Sigma}_\omega) + \mathbf{E} \quad (7)$$

where

$$\begin{aligned} \mathbf{C} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n m_{1ij} \mathbf{B}_{j,k}^T \otimes \mathbf{A}_{k,i} \\ \mathbf{D} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n m_{2ij} \mathbf{B}_{j,k}^T \otimes \mathbf{A}_{k,i} \end{aligned}$$

and

$$\mathbf{E} = \text{vec}(\text{btr}(\mathbf{B})).$$

Upon substituting $E[\text{btr}(\mathbf{Q})] = \text{btr}(\mathbf{Q})$, $\boldsymbol{\Sigma}_\beta = \hat{\boldsymbol{\Sigma}}_\beta$ and $\boldsymbol{\Sigma}_\omega = \hat{\boldsymbol{\Sigma}}_\omega$ in equation (7), the method of moments gives one estimating equation in the vectorized form of two unknown covariance matrices.

4.3.2. Evaluating $E[\text{btr}(\mathbf{Q}_d)]$ and deriving the second estimating equation. Model (5) depends upon one unknown covariance matrix, $\boldsymbol{\Sigma}_\beta$. The intuition is that, upon using all D of the \mathbf{Q}_d matrices in (6) and the method of moments to estimate $\boldsymbol{\Sigma}_\beta$, we will then be able to estimate the other unknown covariance matrix $\boldsymbol{\Sigma}_\omega$ using the first estimating equation. We define design specific $\mathbf{A}_d = (\mathbf{I}_{pn_d} - \mathbf{H}_d)^T \mathbf{W}_d$ and $\mathbf{B}_d = (\mathbf{I}_{pn_d} - \mathbf{H}_d)^T \mathbf{R}_d$, where \mathbf{A}_d and \mathbf{B}_d are known $pn_d \times pn_d$

matrices. We also divide the matrices \mathbf{A} and \mathbf{B} into n_d^2 blocks of $p \times p$ matrices, and write $\mathbf{A}_{d,i,j}$ and $\mathbf{B}_{d,i,j}$, $i, j = 1, \dots, n_d$, to mean the i th by j th blocks of \mathbf{A}_d and \mathbf{B}_d respectively. In the supplementary materials we show that

$$\text{vec} \left(E \left[\sum_{d=1}^D \text{btr}(\mathbf{Q}_d) \right] \right) = \left(\sum_{d=1}^D \mathbf{C}_d \right) \text{vec}(\boldsymbol{\Sigma}_\beta) + \sum_{d=1}^D \mathbf{E}_d \quad (8)$$

where

$$\mathbf{C}_d = \sum_{i=1}^{n_d} \sum_{j=1}^{n_d} \sum_{k=1}^{n_d} m_{1ij}^d \mathbf{B}_{d,j,k}^T \otimes \mathbf{A}_{d,k,i}$$

and

$$\mathbf{E}_d = \text{vec}(\text{btr}(\mathbf{B}_d)).$$

Upon substituting $E[\sum_{d=1}^D \text{btr}(\mathbf{Q}_d)] = \sum_{d=1}^D \text{btr}(\mathbf{Q}_d)$ and $\boldsymbol{\Sigma}_\beta = \hat{\boldsymbol{\Sigma}}_\beta$ in (8), we obtain a second estimating equation from the method of moments.

4.4. Solving the Estimating Equations and Performing Inference

We solve the estimating equation resulting from (8) for $\text{vec}(\hat{\boldsymbol{\Sigma}}_\beta)$ and substitute this estimate into the estimating equation resulting from (7) and solve for $\text{vec}(\hat{\boldsymbol{\Sigma}}_\omega)$.

4.4.1. Estimating $\boldsymbol{\Sigma}_\beta$ under the consistent model. Some applied analysts may prefer to assume the consistent model ($\boldsymbol{\Sigma}_\omega = \mathbf{0}$). As in the univariate case (Jackson et al., 2016), we have two possible ways of estimating $\boldsymbol{\Sigma}_\beta$ under the consistent model: we can use the estimating equation resulting from (7) with $\boldsymbol{\Sigma}_\omega = \mathbf{0}$ or the estimating equation resulting from (8) as in the full model. Also as in the univariate case, we suggest the former option because it uses the information made by assuming consistency when estimating $\boldsymbol{\Sigma}_\beta$. However, this first option is valid only under the consistent model.

4.4.2. "Truncating" the estimates of the unknown covariance matrices so that they are symmetric and positive semi-definite. As in the univariate case, there is the problem that the point estimates of the two unknown covariance matrices are not necessarily positive semi-definite. The method of moments does not even initially enforce the constraint that the point estimates of the unknown covariance matrices are symmetrical (Chen et al., 2012; Jackson et al., 2013). We produce symmetric estimators corresponding to an estimated covariance matrix of $\hat{\boldsymbol{\Sigma}}$ as $(\hat{\boldsymbol{\Sigma}}^T + \hat{\boldsymbol{\Sigma}})/2$ (Chen et al., 2012; Jackson et al., 2013). This also corresponds to taking the average of estimates that result from our \mathbf{Q} and \mathbf{Q}_d matrices and their transposes (Jackson et al., 2013). We then write these symmetric estimators in terms of their spectral decomposition (Chen et al., 2012; Jackson et al., 2013) and truncate any negative eigenvalues to zero to provide the final symmetric positive semi-definite estimated covariance matrices. Specifically, we define the truncated estimate corresponding to the symmetrical $\hat{\boldsymbol{\Sigma}}$ as

$\hat{\Sigma}^+ = \sum_{i=1}^p \max(0, \lambda_i) \mathbf{e}_i \mathbf{e}_i^T$, where λ_i is the i th eigenvalue of the symmetric $\hat{\Sigma}$ and \mathbf{e}_i is the corresponding normalized eigenvector.

4.4.3. Inference for δ . Inference for δ then proceeds as a weighted regression where all weights are treated as fixed and known. Writing $\hat{\mathbf{V}}$ as the estimated variance of \mathbf{Y} in (3), in the absence of missing outcome data we have $\hat{\delta} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{Y}$ where $\text{Var}(\hat{\delta}) = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1}$. In the presence of missing data we can, under our missing completely at random assumption, apply these standard formulae for weighted regression to the observed outcomes. Alternatively and equivalently, we can impute the missing outcome data in \mathbf{Y} with an arbitrary value and replace $\hat{\mathbf{V}}^{-1}$ with the precision matrix corresponding to $\hat{\mathbf{V}}$, calculated in the way explained for \mathbf{S} in Section 4.1 (Jackson et al., 2011). Approximate confidence intervals and hypothesis tests for all basic parameters for all outcomes then immediately follow by taking $\hat{\delta}$ to be approximately normally distributed. Inferences for functional parameters follow by taking appropriate linear combinations of $\hat{\delta}$.

4.5. Special Cases of the Estimation Procedure

In the supplementary materials, we show that the proposed method reduces to two previous methods in special cases. If all studies are two arm studies and consistency is assumed then the proposed method reduces to the matrix based method for multivariate meta-regression (Jackson et al., 2013). The proposed multivariate method reduces to the univariate DerSimonian and Laird method for network meta-analysis (Jackson et al., 2016) when $p = 1$.

4.6. Model Identification

If the necessary standard matrix inversions resulting from the estimating equations from (7) and (8) cannot be performed then both unknown variance components cannot be identified using the proposed method. A minimum requirement for any multivariate modeling is that the common-effect and consistent model must be identifiable. This means that there must be some information (direct or indirect) about each basic parameter for all outcomes. Two or more studies of the same

design must provide data for all possible pairs of outcomes to identify Σ_β . Two or more studies of different designs must provide data for all possible pairs of outcomes to identify Σ_ω . If these conditions are satisfied then the model will be identifiable. In situations where our model is not identifiable we suggest that simpler models should be considered instead. Possible strategies for this include considering models of lower dimension or the consistent model. In practice, it is highly desirable to have more than the minimum amount of replication required, both within and between designs, so that the model is well identified. We make some pragmatic decisions in the next section for our example to provide sufficient replication within designs, in order to estimate Σ_β with reasonable precision.

5. Example

The methodology developed in this article is now applied to an illustrative example in relapsing remitting multiple sclerosis (RRMS). Multiple sclerosis (MS) is an inflammatory disease of the brain and spinal cord and RRMS is a common type of MS. The effectiveness of a new treatment is typically measured to assess its impact on relapse rate and odds of disease progression. Magnetic Resonance Imaging (MRI) allows measurement of the number of new or enlarging lesions in the brain. Three outcomes are included in our analyses, so that $p = 3$ in the full three-dimensional network meta-analysis. These three outcomes are: (i) the log rate ratio of new or enlarging MRI lesions; (ii) the log annualized relapse rate ratio; and (iii) log disability progression odds ratio. Relapse is defined as appearance of new, worsening or recurrence of neurological symptoms that can be attributable to MS, accompanied by an increase of a score on the Expanded Disability Status Scale (EDSS) and also functional-systems score(s), lasting at least 24 hours, preceded by neurologic stability for at least 30 days. Disability progression is defined as an increase in EDSS score that was sustained for 12 weeks, with an absence of relapse at the time of assessment. Negative basic parameters indicate that treatments B-F are beneficial compared to treatment A throughout.

Data in this illustrative example were obtained from ten randomized controlled trials of six treatment options

Table 1
Summary of the relapsing remitting multiple sclerosis dataset.

Study	Design	Outcomes
IFNB SG (1)	AB	All three outcomes measured
IFNB SG (2)	AB	All three outcomes measured
Jacobs/Simon	AC	All three outcomes measured
PRISMS (1)	AC	All three outcomes measured
PRISMS (2)	AC	All three outcomes measured
Johnson	AD	Relapse rate and disability progression only
Durelli	BC	Relapse rate and disability progression only
O'Connor (1)	BD	Relapse rate and disability progression only
O'Connor (2)	BD	Relapse rate and disability progression only
Mikol	CD	All three outcomes measured
FREEDOMS 1	AEF	All three outcomes measured
FREEDOMS 2	AEF	All three outcomes measured
TRANSFORMS	CEF	All three outcomes measured

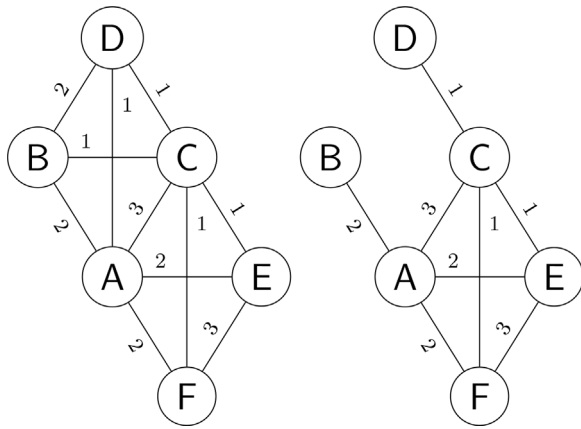


Figure 1. Network diagram for RRMS dataset. A – placebo, B – interferon beta-1b, C – interferon beta-1a, D – glatiramer, E – fingolimod 0.5 mg, F – fingolimod 1.25 mg. Left-hand-side network corresponds to studies reporting the log annualized relapse rate ratio and log disability progression odds ratio (y_2 and y_3) for which data are complete. The right-hand-side network corresponds to studies reporting the log rate ratio of new or enlarging MRI lesions (y_1 which is not reported in four studies). The numbers shown on the network edges are the number of direct comparisons of each pair of treatments; the absence of an edge indicates that there is no direct comparison. Three of the thirteen studies are three arm trials which are each taken to provide three direct comparisons (a direct comparison between each treatment pair). Hence, there are 19 direct comparisons in the left-hand-side network where there is no missing data.

(coded in the network data as treatments A–F); placebo (A), interferon beta-1b (B), interferon beta-1a (C), glatiramer (D), and two doses of fingolimod; 0.5 mg (E) and 1.25 mg (F). Three trials of fingolimod were three-arm (two doses and a control) and are included as three-arm studies. Three trials of interferon beta (one 1a and two 1b) were three-arm (also two doses and a control), and these were included as separate two-arm trials (each dose against the control, with the number of participants in each control arm halved). This ignores the differences in doses of interferon beta and was a pragmatic decision to help provide an identifiable network. Briefly, in this example there is very little replication within designs, so that identifying Σ_β well is very difficult without making pragmatic decisions such as this. Sormani et al. (2010) also treat these particular studies as two separate studies in this way, which helps them to identify their meta-regression models. Treating these three studies as separate two-arm trials means that the data are analyzed as being from thirteen studies and a summary of the resulting data structure is shown in Table 1. There are eight different designs in Table 1 and so there is relatively little replication within designs, even when including three of the three-arm studies as separate two arm studies. Full details of the dataset that are relevant to this article are described in the supplementary materials and see also Bujkiewicz et al. (2016). Figure 1 provides network diagrams that show the number of comparisons between each pair of treatments on the edges. In these diagrams, the three arm studies (Table 1) are taken to contribute three comparisons, for example, the CEF study contributes CE, CF, and EF comparisons. Two estimates of treatment effect from this study contribute to analyses however because C is taken as the baseline; the study's estimated EF treatment effect contains no additional information once its CE and CF contrasts are included in the analysis.

Table 2
Treatment effect estimates of each treatment relative to the reference treatment A (placebo).

Model	AB	AC	Estimate (se) AD	AE	AF
<i>MRI (y_1)</i>					
Univariate (y_1)	−0.95 (0.39)	−1.00 (0.21)	−0.68 (0.50)	−1.38 (0.26)	−1.52 (0.26)
Bivariate (y_1, y_2)	−0.94 (0.39)	−1.00 (0.21)	−0.68 (0.50)	−1.39 (0.26)	−1.53 (0.26)
Bivariate (y_1, y_3)	−0.96 (0.39)	−0.98 (0.22)	−0.66 (0.50)	−1.38 (0.26)	−1.51 (0.26)
trivariate (y_1, y_2, y_3)	−0.96 (0.39)	−0.97 (0.22)	−0.67 (0.50)	−1.38 (0.26)	−1.51 (0.26)
<i>Relapse rate (y_2)</i>					
Univariate (y_2)	−0.35 (0.10)	−0.25 (0.09)	−0.34 (0.11)	−0.81 (0.12)	−0.78 (0.12)
Bivariate (y_1, y_2)	−0.35 (0.10)	−0.25 (0.09)	−0.34 (0.11)	−0.81 (0.12)	−0.78 (0.12)
Bivariate (y_2, y_3)	−0.36 (0.11)	−0.23 (0.10)	−0.33 (0.12)	−0.80 (0.13)	−0.77 (0.13)
trivariate (y_1, y_2, y_3)	−0.36 (0.11)	−0.23 (0.10)	−0.33 (0.12)	−0.80 (0.13)	−0.77 (0.13)
<i>Disability progression (y_3)</i>					
Univariate (y_3)	−0.46 (0.25)	−0.11 (0.21)	−0.42 (0.25)	−0.33 (0.25)	−0.37 (0.24)
Bivariate (y_2, y_3)	−0.47 (0.25)	−0.10 (0.21)	−0.43 (0.25)	−0.37 (0.25)	−0.37 (0.25)
Bivariate (y_1, y_3)	−0.46 (0.25)	−0.11 (0.21)	−0.42 (0.25)	−0.34 (0.25)	−0.38 (0.25)
Trivariate (y_1, y_2, y_3)	−0.47 (0.25)	−0.10 (0.21)	−0.43 (0.25)	−0.37 (0.25)	−0.37 (0.25)

Table 3
Inconsistency and heterogeneity covariance matrices estimates.

Model	$\Sigma_{\omega 11}$	$\Sigma_{\omega 12}$	$\Sigma_{\omega 13}$	$\Sigma_{\omega 22}$	$\Sigma_{\omega 23}$	$\Sigma_{\omega 33}$
Univariate (y_1)	0.0000					
Univariate (y_2)				0.0115		
Univariate (y_3)						0.0713
Bivariate (y_1, y_2)	0.0002	0.0017		0.0125		
Bivariate (y_1, y_3)	0.0018		0.0116			0.0741
Bivariate (y_2, y_3)				0.0161	0.0344	0.0735
Trivariate (y_1, y_2, y_3)	0.0027	0.0066	0.0143	0.0161	0.0349	0.0756
	$\Sigma_{\beta 11}$	$\Sigma_{\beta 12}$	$\Sigma_{\beta 13}$	$\Sigma_{\beta 22}$	$\Sigma_{\beta 23}$	$\Sigma_{\beta 33}$
Univariate (y_1)	0.1508					
Univariate (y_2)				0.0043		
Univariate (y_3)						0.0000
Bivariate (y_1, y_2)	0.1523	-0.0110		0.0047		
Bivariate (y_1, y_3)	0.1526		-0.0191			0.0024
Bivariate (y_2, y_3)				0.0061	0.0015	0.0004
Trivariate (y_1, y_2, y_3)	0.1538	-0.0116	-0.0195	0.0059	0.0024	0.0027

Table 2 shows the estimates of the basic parameters (treatment effects relative to the reference treatment, placebo) obtained from univariate network meta-analyses, bivariate analyses for all three combinations of pairs of outcomes and the trivariate analysis. The results are similar across all analyses, and conclusions from univariate and multivariate analyses are the same. This is disappointing because multivariate analyses have not resulted in more precise inference. The entries of $\hat{\Sigma}_{\beta}$ and $\hat{\Sigma}_{\omega}$ are shown in Table 3. The positive estimates obtained for the unknown variance components suggest that this example exhibits some between-study heterogeneity and inconsistency. In order to assess the impact of the unknown variance components, we also fitted the consistent model and the common-effect and consistent model (results not shown) using all three outcomes ($p = 3$). On average, the standard errors of the fifteen basic parameters from the full model are 35% greater (range: 13–84%) than those from the consistent model, which in turn are 58% (range: 8–128%) greater than those from the common-effect and consistent model. Both the between-study heterogeneity and inconsistency have notable impact.

The multivariate analysis adds to the univariate analyses in two main ways. Firstly, the finding that the multivariate analysis is in good agreement with the univariate analyses is a particularly important finding for treatment effects on MRI where a substantial proportion of data were missing. It has been demonstrated by Kirkham et al. (2012) that a multivariate approach to meta-analysis can help obtain more accurate estimates in the presence of outcome reporting bias. Hence, the multivariate analysis reduces concerns that this univariate analysis is affected by reporting bias. Secondly joint inferences for all three outcomes are possible under the multivariate model. For example, and as we might anticipate, in our example the estimated log annualized relapse rate ratios and log disability progression odds ratios are highly positively correlated; from $\text{Var}(\hat{\delta})$ in our three-dimensional multivariate meta-analysis, the correlations between the five pairs of estimated basic parameters for these two outcomes are all

between 0.63 and 0.75. Medical decision making based jointly on these two outcomes should take this high positive correlation into account, and this is only possible by using a multivariate approach. For example, a formal decision analysis involving these two outcomes should be based on their joint distribution rather than their two marginal distributions. In the supplementary materials, we perform a simulation study to further explore how the proposed methodology performs.

6. Discussion

We have proposed a new model for dealing with both multiple treatment contrasts and multiple outcomes, to provide a framework for conducting multivariate network meta-analysis. By using a matrix-based method of moments estimator, our methodology naturally builds on previous work (such as the well-known DerSimonian and Laird approach) and is computationally very fast, relative to other potential estimation approaches such as REML or MCMC; this is especially the case in very high dimensions and so our methodology is particularly advantageous for ambitious analyses of this type. The main disadvantage is that, as a necessary consequence of its semi-parametric nature, the method of moments is not based on sufficient statistics and so is not fully efficient. The loss in efficiency relative to maximum likelihood estimation awaits investigation but we anticipate that this will be less serious for inferences about the average effects than the unknown variance components. Furthermore, the within-study normal approximations used in our model are not necessarily very accurate even in moderately sized studies.

Since our analysis uses a general design matrix, the modeling may easily be extended by adding study level covariates to describe and fit multivariate network meta-regressions. In the network meta-analysis setting, these regressions have the potential to explain the reasons for inconsistency and model multiple dose level responses. Our method of moments estimation can be combined with approaches that “inflate”

confidence intervals from a frequentist random effects meta-analysis (Hartung and Knapp, 2001; Jackson and Riley, 2014).

In conclusion, we have developed a new model and estimation method for multivariate network meta-analysis, which can describe multiple treatments and multiple correlated outcomes. An R function is available in the web supplementary materials that implements the proposed methodology.

7. Supplementary Materials

Web appendices referenced in Sections 2, 4, and 5 are available with this article at the *Biometrics* website on Wiley Online Library. Computing codes are also available.

ACKNOWLEDGEMENTS

DJ, IRW, and ML are (or were) employed by the UK Medical Research Council [Unit Programme number U105260558]. SB was supported by the Medical Research Council (MRC) Methodology Research Programme [New Investigator Research Grant MR/L009854/1].

REFERENCES

- Achana, F. A., Cooper, N. J., Bujkiewicz, S., Hubbard, S. J., Kendrick, D., Jones, D. R., et al. (2014). Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Medical Research Methodology* **14**, 92.
- Bujkiewicz, S., Thompson, J. R., Riley, R. D., and Abrams, K. R. (2016). Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. *Statistics in Medicine* **35**, 1063–1089.
- Chen, H., Manning, A. K., and Dupuis, J. (2012). A method of moments estimator for random effect multivariate meta-analysis. *Biometrics* **68**, 1278–1284.
- Dersimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.
- Efthimiou, O., Mavridis, D., Cipriani, A., Leucht, S., Bago, P., and Salanti, G. (2014). An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Statistics in Medicine* **33**, 2275–2287.
- Efthimiou, O., Mavridis, D., Riley, R. D., Cipriani, A., and Salanti, G. (2015). Joint synthesis of multiple correlated outcomes in networks of interventions. *Biostatistics* **16**, 84–97.
- Hartung, J. and Knapp, G. (2001). On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine* **20**, 1771–1782.
- Henderson, H. V. and Searle, S. R. (1981). The vec-permutation matrix, the vec operator and Kronecker products: A review. *Linear and Multilinear Algebra* **9**, 271–288.
- Higgins, J. P. T., Jackson, D., Barrett, J. K., Lu, G., Ades, A. E., and White, I. R. (2012). Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods* **3**, 98–110.
- Hong, H., Fu, H., Price, K. L., and Carlin, B. P. (2015). Incorporation of individual-patient data in network meta-analysis for multiple continuous endpoints, with application to diabetes treatment. *Statistics in Medicine* **34**, 2794–2819.
- Hong, H., Chu, H., Zhang, J., and Carlin, B. P. (2016). A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Research Synthesis Methods* **7**, 6–22.
- Jackson, D., Riley, R., and White, I. R. (2011). Multivariate meta-analysis: potential and promise (with discussion). *Statistics in Medicine* **30**, 2481–2510.
- Jackson, D., White, I. R., and Riley, R. D. (2013). A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. *Biometrical Journal* **55**, 231–245.
- Jackson, D. and Riley, R. (2014). A refined method for multivariate meta-analysis and meta-regression. *Statistics in Medicine* **33**, 541–554.
- Jackson, D., Law, M., Barrett, J. K., Turner, R., Higgins, J. P. T., Salanti, G., et al. (2016). Extending DerSimonian and Laird's methodology to perform network meta-analyses with random inconsistency effects. *Statistics in Medicine* **35**, 819–839.
- Kirkham, J. J., Riley, R. D., and Williamson, P. R. (2012). A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine* **31**, 2179–2195.
- Law, M., Jackson, D., Turner, R., Rhodes, K., and Viechtbauer, W. (2016). Two new methods to fit models for network meta-analysis with random inconsistency effects. *BMC Medical Research Methodology* **16**, 87.
- Lu, G. and Ades, A. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* **23**, 3105–3124.
- Lu, G. and Ades, A. (2006). Assessing evidence consistency in mixed treatment comparisons. *Journal of the American Statistical Association* **101**, 447–459.
- Nikolakopoulou, A., Chaimani, A., Veroniki, A., Vasiliadis, H. S., Schmid, C. H., and Salanti, G. (2014). Characteristics of networks of interventions: A description of a database of 186 published networks. *PLOS ONE* **9**, 1: e86754.
- Piepho, H. P., Williams, E. R., and Madden, L. V. (2012). The use of two-way linear mixed models in multitreatment meta-analysis. *Biometrics* **68**, 1269–1277.
- Riley, R. D., Thompson, J. R., and Abrams, K. R. (2008). An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* **9**, 172–186.
- Sormani, M. P., Bonzano, L., Roccatagliata, L., Mancardi, G. L., Uccelli, A., and Bruzzi, P. (2010). Surrogate endpoints for EDSS worsening in multiple sclerosis. A meta-analytic approach. *Neurology* **75**, 302–309.
- Seaman, S., Galati, J., Jackson, D., and Carlin, J. (2013). What Is Meant by “Missing at Random?” *Statistical Science* **28**, 257–268.
- Searle, S. R. (1971). *Linear Models*. New York: Wiley.
- Veroniki, A., Vasiliadis, H. S., Higgins, J. P., and Salanti, G. (2013). Evaluation of inconsistency in networks of interventions. *International Journal of Clinical Epidemiology* **42**, 332–345.
- Wei, Y. and Higgins, J. P. T. (2013). Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in Medicine* **32**, 1191–1205.

Received June 2016. Revised June 2017. Accepted June 2017.